ELSEVIER

Stochastic Processes and their Applications 95 (2001) 133–149

stochastic processes and their applications

www.elsevier.com/locate/spa

Forward and backward diffusion approximations for haploid exchangeable population models

M. Möhle

Fachbereich Mathematik, Johannes Gutenberg-Universität Mainz, Saarstraße 21, 55099 Mainz, Germany

Received 25 July 2000; received in revised form 7 February 2001; accepted 5 March 2001

Abstract

The class of haploid population models with non-overlapping generations and fixed population size N is considered such that the family sizes $\nu_1,\dots,\nu_N$ within a generation are exchangeable random variables. A criterion for weak convergence in the Skorohod sense is established for a properly time- and space-scaled process counting the number of descendants forward in time. The generator A of the limit process X is constructed using the joint moments of the offspring variables $\nu_1,\dots,\nu_N$. In particular, the Wright–Fisher diffusion with generator $Af(x)=\frac{1}{2}x(1-x)f''(x)$ appears in the limit as the population size N tends to infinity if and only if the condition $\lim_{N\to\infty} E((\nu_1-1)^3)/(N\operatorname{Var}(\nu_1))=0$ is satisfied.

Using the concept of duality, these convergence results are compared with the limit theorems known for the coalescent processes with simultaneous and multiple collisions arising when the models are considered backward in time. In particular the Wright–Fisher diffusion appears forward in time if and only if the Kingman coalescent appears backward in time as N tends to infinity. A commutative diagram leads to a full understanding of the model considered forward and backward in time for finite population size and in the limit as N tends to infinity.

MSC: primary 92D25, 60J70; secondary 92D15, 60F17

Keywords: Ancestors; Coalescent; Descendants; Diffusion approximation; Duality; Exchangeability; Hille–Yosida theorem; Neutrality; Population genetics; Weak convergence; Wright–Fisher diffusion

1. Introduction

In 1974 Cannings introduced a class of haploid population models with non-overlapping generations and fixed population size $N\in\mathbb{N}:=\{1,2,\dots\}$. Each model in this class is characterized by exchangeable random variables $\nu_1,\dots,\nu_N$, where ν_i denotes the number of offspring of the ith individual. As the population size is assumed to be fixed the equation $\nu_1+\cdots+\nu_N=N$ has to be satisfied.

Sample a certain number of individuals from the current generation and for $r\in\mathbb{N}_0:=\{0,1,2,\dots\}$ let $\mathscr{X}_r$ denote the (random) number of descendants of the sampled

E-mail address: moehle@mathematik.uni-mainz.de (M. Möhle).

0304-4149/01/$ - see front matter
PII: S0304-4149(01)00093-X

individuals r generations forward in time. It is well known (Cannings, 1974, 1975) that $\mathscr{X} := (\mathscr{X}_r)_{r\in\mathbb{N}_0}$ is a time-homogeneous Markov chain with state space $\{0,\dots,N\}$ and transition probabilities $\pi_{ij} := P(\mathscr{X}_{r+1} = j \mid \mathscr{X}_r = i)$ given by

$$\pi_{ij} = P(v_1 + \cdots + v_i = j), \tag{1}$$

$i,j\in\{0,\dots,N\}$. The process $\mathscr{X}$ is called the *forward process* or *descendant process*. The most celebrated example is the so-called Wright–Fisher model formulated implicitly by Fisher (1922) and explicitly by Wright (1931), where $(v_1,\dots,v_N)$ has a symmetric multinomial distribution. For the Wright–Fisher model it is known that the time- and space-scaled forward process $(\mathscr{X}_{[Nt]}/N)_{t\geqslant 0}$ converges to a limit diffusion process on $E := [0,1]$ as the population size N tends to infinity. This process, called the *Wright–Fisher diffusion* (with no mutation or selection), has a generator given by

$$Af(x) = \tfrac{1}{2}x(1-x)f''(x), \tag{2}$$

$x\in E$, $f\in C^2(E)$. The Wright–Fisher diffusion and similar diffusion approximations have been obtained and studied by various authors, for example by Crow and Kimura (1970), Ethier and Nagylaki (1980), Norman (1972) and Trotter (1958). Sato (1976) studied the more general multi-allelic case. The novelty in this paper is that the complete class of models with exchangeable reproduction is treated simultaneously, leading to a full classification of all possible limit processes for the case of exchangeable reproduction. A criterion is presented (Theorem 2.3) which ensures that the process

$$\left(\frac{\mathscr{X}_{[t/c_N]}}{N}\right)_{t\geqslant 0} \tag{3}$$

converges in $D_E([0,\infty))$, i.e. in the Skorohod sense, to a limit process $X := (X_t)_{t\geqslant 0}$ on E, where $D_E([0,\infty))$ denotes the set of all E-valued functions on $[0,\infty)$ which are right continuous and have left limits. Here the time has to be measured in units of $[1/c_N]$ generations, where c_N denotes the so-called coalescence probability defined precisely in (7). The possible limit processes X are characterized via their generators. The criterion is based on an analysis of the functions

$$\Phi_j(k_1,\dots,k_j) := \Phi_j^{(N)}(k_1,\dots,k_j) := \frac{(N)_j}{(N)_k}E((v_1)_{k_1}\cdots(v_j)_{k_j}), \tag{4}$$

$j,k_1,\dots,k_j\in\mathbb{N}$ with $k := k_1+\cdots+k_j \leqslant N$, where $(x)_k := x(x-1)\cdots(x-k+1)$ for $x\in\mathbb{R}$ and $k\in\mathbb{N}_0$. These quantities are well known from ancestral population genetics (Kingman, 1982b; Möhle, 1998). More precisely, $\Phi_j(k_1,\dots,k_j)$ is the probability that k children, chosen randomly from some generation, have exactly j ancestors (parents) one generation backward in time, k_1 of them descended from one parent, k_2 from another parent and so on.

As a corollary it is shown (Theorem 3.3) that the Wright–Fisher diffusion with generator (2) appears in the limit as N tends to infinity if and only if the condition

$$\lim_{N\to\infty}\frac{E((v_1-1)^3)}{N\,\mathrm{Var}(v_1)} = 0 \tag{5}$$

or equivalently (see Lemma 3.2)

$$\phi_1(3) := \lim_{N\to\infty} \frac{\Phi_1^{(N)}(3)}{c_N} = \lim_{N\to\infty} \frac{E((v_1)_3)}{N\,E((v_1)_2)} = 0 \tag{6}$$

is satisfied, where

$$c_N := \Phi_1^{(N)}(2) = \frac{E((v_1)_2)}{N-1} = P(2 \text{ children have same parent}) \tag{7}$$

denotes the so-called *coalescence probability*, i.e. the probability that two individuals, chosen randomly without replacement from some generation, have a common ancestor one generation backward in time. Condition (6) says that

$$P(3 \text{ children have same parent}) = \Phi_1^{(N)}(3) \sim o(c_N)$$

or in other words that triple mergers of ancestral lines are asymptotically negligible in comparison with binary mergers.

While the probabilities (4) are the proper quantities to analyse the model backward in time, it turns out that for an analysis of the forward process $\mathscr{X}$ it is (mainly for computational convenience) more helpful to consider a kind of centered version of the functions (4) defined via

$$\Psi_j(k_1,\dots,k_j) := \Psi_j^{(N)}(k_1,\dots,k_j) := \frac{(N)_j}{(N)_k} E((v_1-1)^{k_1}\cdots(v_j-1)^{k_j}), \tag{8}$$

$j,k_1,\dots,k_j \in \mathbb{N}$ with $k := k_1+\cdots+k_j \leqslant N$. The exchangeability of the offspring variables $v_1,\dots,v_N$ and $v_1+\cdots+v_N = N$ ensure that

$$\begin{aligned}
&(N-j)E((v_1-1)^{k_1}\cdots(v_j-1)^{k_j}(v_{j+1}-1))\\
&\quad = E((v_1-1)^{k_1}\cdots(v_j-1)^{k_j}((v_{j+1}-1)+\cdots+(v_N-1)))\\
&\quad = E((v_1-1)^{k_1}\cdots(v_j-1)^{k_j}(-(v_1-1)-\cdots-(v_j-1)))\\
&\quad = -\sum_{i=1}^{j} E((v_1-1)^{k_1}\cdots(v_i-1)^{k_i+1}\cdots(v_j-1)^{k_j}).
\end{aligned}$$

Multiplication of both sides by $(N)_j/(N)_{k+1}$ leads to the recursion

$$\Psi_{j+1}(k_1,\dots,k_j,1) = -\sum_{i=1}^{j} \Psi_j(k_1,\dots,k_{i-1},k_i+1,k_{i+1},\dots,k_j), \tag{9}$$

$j,k_1,\dots,k_j \in \mathbb{N}$ with $k+1 \leqslant N$. Note that the coalescence probability (7) is expressed in terms of the functions (8) via

$$c_N = \Psi_1^{(N)}(2) = \frac{E((v_1-1)^2)}{N-1} = \frac{\mathrm{Var}(v_1)}{N-1}. \tag{10}$$

The coalescence probability is of fundamental interest in population genetics as $N_\mathrm{e} := [1/c_N]$ is the proper time scale for processes defined forward in time and also for processes defined backward in time (see for example, Möhle (2000b)) in order to reach convergence as the population size N tends to infinity. N_e is called the *effective population size* or *effective population number* (Crow and Kimura, 1970). It is assumed that $c_N > 0$ or equivalently the trivial reproduction law $P(v_1 = \cdots = v_N = 1) = 1$ is excluded.

Section 2 presents the main convergence theorem. Section 3 considers the case when the limit process X is the Wright–Fisher diffusion. In Section 4 the concept of duality is used to compare the model forward and backward in time and to connect the limit process X with ancestral processes known from the coalescent theory going back to Kingman (1982a,b,c). The paper finishes with examples presented in Section 5.

2. The limit theorem

This section presents the main convergence theorem (Theorem 2.3). Its proof is based on an application of semigroup theory well described in Ethier and Kurtz (1986). The theorem provides a criterion under which the process (3) converges as the population size N tends to infinity in the Skorohod sense to a limit process X characterized via a certain generator A. The criterion is based on the existence of the limits

$$\psi_j(k_1,\dots,k_j) := \lim_{N\to\infty} \frac{\Psi_j^{(N)}(k_1,\dots,k_j)}{c_N} = \lim_{N\to\infty} \frac{E((v_1-1)^{k_1}\cdots(v_j-1)^{k_j})}{N^{k_1+\cdots+k_j-j}c_N}, \tag{11}$$

$j,k_1,\dots,k_j\in\mathbb{N}$. Note that if the limits (11) exist for all $j\in\mathbb{N}$ and $k_1\geqslant\cdots\geqslant k_j\geqslant 2$, then the recursion (9) ensures that the limits (11) exist for all $j,k_1,\dots,k_j\in\mathbb{N}$ and the limits (11) satisfy the same recursion. The following lemma provides the existence of certain polynomials $a_k(x)$, which play a crucial role in constructing the generator A of the limit process X. For $i\in\{0,\dots,N\}$ let

$$C_i := \sum_{j=1}^{i} v_j \tag{12}$$

denote the number of children of the parents $1,\dots,i$.

Lemma 2.1. *If the limits* (11) *exist for all* $j\in\mathbb{N}$ *and* $k_1\geqslant\cdots\geqslant k_j\geqslant 2$ *then also the limits*

$$a_k(x) := \lim_{N\to\infty}\frac{1}{c_N}E\left(\left(\frac{1}{N}\sum_{j=1}^{[Nx]}(v_j-1)\right)^k\right) = \lim_{N\to\infty}\frac{1}{c_N}E\left(\left(\frac{C_{[Nx]}}{N}-\frac{[Nx]}{N}\right)^k\right) \tag{13}$$

exist for all $k\in\mathbb{N}$ *and* $x\in E$, *where for fixed* k *this convergence holds uniformly in* $x\in E$.

Remark. Note that (13) is equivalent to

$$E\left(\left(\frac{C_{[Nx]}}{N}-\frac{E(C_{[Nx]})}{N}\right)^k\right) \sim c_N\, a_k(x),$$

saying that the kth centered moment of the relative number of children of $[Nx]$ parents is asymptotically equal to the coalescence probability c_N times a constant, namely $a_k(x)$.

Proof of Lemma 2.1. Fix $k \in \mathbb{N}$. For $i \in \{0,\dots,N\}$ consider

$$
\begin{aligned}
E((C_i-i)^k) &= E\left(\left(\sum_{j=1}^{i}(v_j-1)\right)^k\right)\\
&= \sum_{\substack{k_1,\dots,k_i\in\mathbb{N}_0\\ k_1+\cdots+k_i=k}} \frac{k!}{k_1!\cdots k_i!} E((v_1-1)^{k_1}\cdots(v_i-1)^{k_i})\\
&= \sum_{j=1}^{k}\binom{i}{j} \sum_{\substack{k_1,\dots,k_j\in\mathbb{N}\\ k_1+\cdots+k_j=k}} \frac{k!}{k_1!\cdots k_j!} E((v_1-1)^{k_1}\cdots(v_j-1)^{k_j}).
\end{aligned}
$$

Note that this is in particular satisfied for $i=0$ as empty sums are defined as zero. A multiplication of the last chain of equations with $1/(N)_k$ $(N \geqslant k)$ leads to

$$
\begin{aligned}
\frac{E((C_i-i)^k)}{(N)_k} &= \sum_{j=1}^{k}\binom{i}{j} \sum_{\substack{k_1,\dots,k_j\in\mathbb{N}\\ k_1+\cdots+k_j=k}} \frac{k!}{k_1!\cdots k_j!} \frac{E((v_1-1)^{k_1}\cdots(v_j-1)^{k_j})}{(N)_k}\\
&= \sum_{j=1}^{k}\frac{\binom{i}{j}}{(N)_j} \sum_{\substack{k_1,\dots,k_j\in\mathbb{N}\\ k_1+\cdots+k_j=k}} \frac{k!}{k_1!\cdots k_j!} \Psi_j^{(N)}(k_1,\dots,k_j),
\end{aligned}
\tag{14}
$$

where $\binom{i}{j} := 0$ for $j > i$. Now multiply with $1/c_N$, choose $i := [Nx]$ with $x \in E$ and let N tend to infinity in the last chain of equalities to verify that the limit (13) exists and is a polynomial of the form

$$
a_k(x) = \sum_{j=1}^{k}\frac{x^j}{j!} \sum_{\substack{k_1,\dots,k_j\in\mathbb{N}\\ k_1+\cdots+k_j=k}} \frac{k!}{k_1!\cdots k_j!} \psi_j(k_1,\dots,k_j),
\tag{15}
$$

where $\psi_j(k_1,\dots,k_j)$ is given by (11). Note that for fixed $k \in \mathbb{N}$ the convergence is uniformly in $x \in [0,1]$ as $\binom{[Nx]}{j}/(N)_j$ converges uniformly to $x^j/j!$ on E and as all the sums on the right hand side of the appearing equations are finite. □

Remarks.

1. In this remark the polynomials $a_1(x)$, $a_2(x)$ and $a_3(x)$ are calculated. Note that $\Psi_1(1) = E(v_1-1) = 0$, i.e. $a_1(x) = x\psi_1(1) = 0$. The recursion (9) yields $\Psi_2(1,1) = -\Psi_1(2) = -c_N$. Thus from (14) with $k=2$ it follows that:

$$
\frac{E((C_i-i)^2)}{(N)_2} = \frac{i}{N}\Psi_1(2) + \frac{(i)_2}{(N)_2}\Psi_2(1,1) = \frac{i(N-i)}{(N)_2}c_N
$$

or equivalently $N^{-2}E((C_i-i)^2)=i/N(1-i/N)c_N$. In particular $a_2(x)=x(1-x)$. The larger k is, the more complicated is the structure of $a_k(x)$. For example for the case $k=3$ it follows from (14) that

$$\frac{E((C_i-i)^3)}{(N)_3}=\frac{i}{N}\Psi_1(3)+3\frac{(i)_2}{(N)_2}\Psi_2(2,1)+\frac{(i)_3}{(N)_3}\Psi_3(1,1,1).$$

The recursion (9) shows that $\Psi_2(2,1)=-\Psi_1(3)$ and that $\Psi_3(1,1,1)=-2\Psi_2(2,1)=2\Psi_1(3)$ and it follows that

$$\begin{aligned}\frac{E((C_i-i)^3)}{(N)_3}&=\left(\frac{i}{N}-3\frac{(i)_2}{(N)_2}+2\frac{(i)_3}{(N)_3}\right)\Psi_1(3)\\&=\frac{i(N-i)(N-2i)}{(N)_3}\Psi_1(3)\end{aligned}$$

or equivalently $N^{-3}E((C_i-i)^3)=i/N(1-i/N)(1-2i/N)\Psi_1(3)$. Thus $a_3(x)=x(1-x)(1-2x)\psi_1(3)$ depends already on the limit $\psi_1(3)$. Note that $a_3(\frac{1}{2})=0$ and that $a_3(x)>0$ for $x\in(0,\frac{1}{2})$ and $a_3(x)<0$ for $x\in(\frac{1}{2},1)$ whenever $\psi_1(3)>0$. If the condition (5) (or equivalently (6)) holds, then we have $a_3(x)=0$ and, moreover, the forthcoming monotonicity in Lemma 3.1 and equivalence of the limits in Lemma 3.2 yield $a_k(x)=0$ for all $k\geqslant 3$. In this case the generator A defined later (see (18)) reduces to the generator (2) of the Wright–Fisher diffusion.

From (13) and $|C_i-i|/N\leqslant 1$ almost surely it follows that $0\leqslant a_{2k+2}(x)\leqslant a_{2k}(x)$ and $|a_{2k+1}(x)|\leqslant a_{2k}(x)$ for all $k\in\mathbb{N}$. In particular $|a_k(x)|\leqslant a_2(x)$ and $a_k(0)=a_k(1)=0$ for all $k\in\mathbb{N}$.

2. For $x\in E$ and $\Gamma\in\mathscr{B}(E)$ define $\mu_N(x,\Gamma):=P(C_{[Nx]}/N\in\Gamma)$. Obviously

$$f(y)=\sum_{k=0}^{n}\frac{f^{(k)}(x)}{k!}(y-x)^k \tag{16}$$

for all polynomials f of degree less than or equal to n. Therefore

$$\begin{aligned}\int_E f(y)\,\mu_N(x,\mathrm{d}y)&=\sum_{k=0}^{n}\frac{f^{(k)}(x)}{k!}\int_E(y-x)^k\,\mu_N(x,\mathrm{d}y)\\&=\sum_{k=0}^{n}\frac{f^{(k)}(x)}{k!}E\left(\left(\frac{C_{[Nx]}}{N}-x\right)^k\right).\end{aligned}$$

If the limits (11) exist for all $j\in\mathbb{N}$ and $k_1\geqslant\cdots\geqslant k_j\geqslant 2$ and if furthermore the limit $c:=\lim_{N\to\infty}c_N$ exists then Lemma 2.1 ensures that

$$\lim_{N\to\infty}\int_E f(y)\,\mu_N(x,\mathrm{d}y)=f(x)+c\sum_{k=1}^{n}\frac{f^{(k)}(x)}{k!}a_k(x).$$

The choice $f(x):=x^n$ shows that the nth moment of the probability measure $\mu_N(x,.)$ converges to

$$x^n+c\sum_{k=1}^{n}\binom{n}{k}x^{n-k}a_k(x). \tag{17}$$

This convergence of moments implies (see Feller, 1971, Chapter 8, Section 1) the weak convergence of the sequence of probability measures $\mu_N(x,.)$ $(N \in \mathbb{N})$ to some limit probability measure $\mu(x,.)$ with moments (17). If $c=0$ then this limit measure is the unit point mass at x.

Let $L := C(E)$ denote the space of all real-valued, continuous (and hence Borel measurable) functions on E. As E is compact all the functions $f \in L$ are bounded and we can supply L with the norm $\|f\| := \sup_{x \in E} |f(x)|$. The subspace of L of all polynomials on E is denoted by D. Assume now that the limits (11) exist for all $j \in \mathbb{N}$ and $k_1 \geqslant \cdots \geqslant k_j \geqslant 2$. Define a linear operator A on L with domain D via

$$Af(x) := \sum_{k=1}^{\infty} \frac{f^{(k)}(x)}{k!} a_k(x), \tag{18}$$

where the $a_k(x)$ are defined via (13). Note that the sum on the right in (18) is finite as long as f is a polynomial, i.e. A is well defined on D. The following lemma presents another representation of the generator A.

Lemma 2.2. *For $x \in E$ and $f \in D$ the formula*

$$\begin{aligned} Af(x) &= \lim_{N\to\infty} \frac{1}{c_N} E\left(f\left(x + \frac{C_{[Nx]} - [Nx]}{N}\right) - f(x)\right) \\ &= \lim_{N\to\infty} \frac{1}{c_N} E\left(f\left(\frac{C_{[Nx]}}{N}\right) - f\left(\frac{[Nx]}{N}\right)\right) \end{aligned}$$

holds, where the C_i, $i \in \{0,\ldots,N\}$ are defined via (12).

Proof. Fix $x \in E$ and define $i := [Nx]$. Applying (16) to the random point $y := x + (C_i - i)/N$ yields

$$c_N^{-1} E\left(f\left(x + \frac{C_i - i}{N}\right) - f(x)\right) = \sum_{k=1}^{n} \frac{f^{(k)}(x)}{k!} c_N^{-1} E\left(\left(\frac{C_i - i}{N}\right)^k\right),$$

where n denotes the degree of the polynomial f. Lemma 2.1 ensures that the last sum, with $i = [Nx]$, converges to $\sum_{k=1}^{n} f^{(k)}(x) a_k(x)/k! = Af(x)$ as N tends to infinity. The second formula follows in a similar way applying (16) to the point $y = C_i/N$ and i/N instead of x. $\square$

In order to apply the Hille–Yosida theorem (Ethier and Kurtz, 1986, p. 165, Theorem 2.2) we have to verify that

(a) D is dense in L,

(b) A satisfies the positive maximum principle, i.e. $Af(x) \leqslant 0$ for all $f \in D$, $x \in E$ with $\sup_{y \in E} f(y) = f(x) \geqslant 0$, and that

(c) the range of $\lambda - A$ is dense in L for some $\lambda > 0$.

Obviously (a) is satisfied (Stone–Weierstraß). The condition (b) follows from the representation of A given in Lemma 2.2. In order to verify (c) let $L_n \subset L$ denote the set of all polynomials with degree less or equal than n, $n \in \mathbb{N}$. Obviously $L_1, L_2, \ldots$ is a sequence of finite-dimensional subspaces of D such that $\bigcup_{n \in \mathbb{N}} L_n = D$. As the a_k

defined via (13) are polynomials of the form (15), i.e. $a_k \in L_k$, we see that $Af \in L_n$ for all $f \in L_n$, i.e. $A : L_n \to L_n$. Obviously $(\lambda - A)(L_n) = L_n$ for all λ not belonging to the set of eigenvalues of $A|L_n$, i.e. for all but at most finitely many $\lambda > 0$. Thus $(\lambda - A)(D) = (\lambda - A)(\bigcup_{n \in \mathbb{N}} L_n) = \bigcup_{n \in \mathbb{N}} L_n = D$ is dense in L for all but at most countable many $\lambda > 0$. In particular (c) is satisfied.

Thus the Hille–Yosida theorem ensures that the closure $\bar{A}$ of A on L is single-valued and generates a strongly continuous, positive, contraction semigroup $\{T(t)\}$ on L. Note that from (c) it follows that D is a core for $\bar{A}$. (see Ethier and Kurtz, 1986, p. 17, Proposition 3.1). For the special case of the Wright–Fisher generator (2) this is for example also mentioned in Ethier and Kurtz (1986, p. 375, Theorem 2.8). Obviously A maps constant functions to the zero function, i.e. A is conservative. Thus $\{T(t)\}$ is a Feller semigroup (see Ethier and Kurtz, 1986, p. 166) and corresponds to a Markov process $X = (X_t)_{t \geqslant 0}$ with sample paths in $D_E([0,\infty))$.

Theorem 2.3. *Assume that $\mathscr{X}_0/N$ converges in distribution to some probability measure ν as N tends to infinity. Assume further that all the limits* (11), $j \in \mathbb{N}$, $k_1 \geqslant \cdots \geqslant k_j \geqslant 2$ *exist and that furthermore the limit $c := \lim_{N\to\infty} c_N$ exists.*

(a) *If $c > 0$ then the process $(\mathscr{X}_r/N)_{r \in \mathbb{N}_0}$ converges weakly as N tends to infinity to a discrete-time Markov process $(X_r)_{r \in \mathbb{N}_0}$ in E with initial distribution ν (i.e. $P(X_0 \in \Gamma) = \nu(\Gamma)$ for all $\Gamma \in \mathscr{B}(E)$) and transition function*

$$P(X_{r+1} \in \Gamma \,|\, X_r = x) = \mu(x, \Gamma),$$

where $\mu(x, .)$, the probability measure uniquely determined via its moments (17), *is the weak limit of the probability measures $\mu_N(x, .)$ defined in the second remark after the proof of Lemma* 2.1.

(b) *If $c = 0$ then the process $(\mathscr{X}_{[t/c_N]}/N)_{t \geqslant 0}$ converges weakly in $D_E([0,\infty))$ as N tends to infinity to a continuous-time Markov process $X = (X_t)_{t \geqslant 0}$ in E with initial distribution ν and generator* (18).

Proof. For $x \in E_N := \{0, 1/N, 2/N, \ldots, 1\}$ and $f \in D$ consider

$$\begin{aligned} T_N f(x) &:= E\left(f\left(\frac{\mathscr{X}_{r+1}}{N}\right) \middle| \frac{\mathscr{X}_r}{N} = x\right) \\ &= \sum_{j=0}^{N} f\left(\frac{j}{N}\right) P(\mathscr{X}_{r+1} = j \,|\, \mathscr{X}_r = i) \\ &= \sum_{j=0}^{N} f\left(\frac{j}{N}\right) P(\nu_1 + \cdots + \nu_i = j) = E\left(f\left(\frac{C_i}{N}\right)\right), \end{aligned}$$

where $i := Nx$ and $C_i := \sum_{j=1}^{i} \nu_j$ for all $i \in \{0, \ldots, N\}$. Note that $C_0 = 0$ and $C_N = N$. If n denotes the degree of the polynomial f then an application of (16) to the random point $y := C_i/N$ leads to

$$T_N f(x) = \sum_{k=0}^{n} \frac{f^{(k)}(x)}{k!} E\left(\left(\frac{C_i}{N} - x\right)^k\right)$$

and hence

$$\left|\frac{T_N f(x) - f(x)}{c_N} - Af(x)\right|$$
$$= \left|\sum_{k=1}^{n} \frac{f^{(k)}(x)}{k!} \frac{E((C_i/N - x)^k)}{c_N} - \sum_{k=1}^{n} \frac{f^{(k)}(x)}{k!} a_k(x)\right|$$
$$\leqslant \sum_{k=1}^{n} \frac{|f^{(k)}(x)|}{k!} \left|\frac{E((C_i/N - x)^k)}{c_N} - a_k(x)\right|.$$

As all the derivatives $f^{(k)}$ are bounded on E, i.e. $\|f^{(k)}\| = \sup_{x \in E} |f^{(k)}(x)| < \infty$ the previous Lemma 2.1 shows that

$$\lim_{N\to\infty} \sup_{x \in E_N} \left|\frac{T_N f(x) - f(x)}{c_N} - Af(x)\right| = 0 \tag{19}$$

for all $f \in D$. Part (b) follows now from Ethier and Kurtz (1986, p. 233, Corollary 8.9).

Part (a) follows also as in this case $T_N f(x) = E(f(C_i/N)) = \int_E f(y)\,\mu_N(x, \mathrm{d}y)$ converges to $\int_E f(y)\,\mu(x, \mathrm{d}y) =: Tf(x)$ for all $f \in L$ as N tends to infinity. From (19) it follows that in this case the generator A in (18) has the form

$$Af(x) = \frac{Tf(x) - f(x)}{c} = \frac{1}{c} \int_E (f(y) - f(x))\,\mu(x, \mathrm{d}y). \qquad \square \tag{20}$$

Remark. Generators having an integral representation (20) are well known (Ethier and Kurtz, 1986, p. 162, Eq. (2.1)). They correspond to Markov jump processes. Note that such generators are bounded, while the generator (2) of the Wright–Fisher diffusion is unbounded.

3. The Wright–Fisher diffusion

The Wright–Fisher diffusion occurs when the limits (13) are equal to zero for all $k \geqslant 3$ such that the generator A in (18) has the form (2). Note that the Wright–Fisher diffusion corresponds to the stochastic differential equation $\mathrm{d}X_t = \sqrt{X_t(1 - X_t)}\,\mathrm{d}B_t$, where $(B_t)_{t \geqslant 0}$ denotes standard Brownian motion. This representation is not needed in this paper. All the proofs are based on the characterization of the Wright–Fisher diffusion via the generator (2).

The main theorem in this section (Theorem 3.3) shows that the Wright–Fisher diffusion appears in the limit as the population size N tends to infinity if and only if the condition (5) or equivalently (6) is satisfied. For the proof of this theorem it turns out to be helpful to compare the functions $\Phi_j(k_1, \ldots, k_j)$ defined in (4) with the functions $\Psi_j(k_1, \ldots, k_j)$ defined in (8).

Lemma 3.1. *The functions* (4) *are monotone in the sense that*

$$\Phi_j(k_1, \ldots, k_j) \leqslant \Phi_l(m_1, \ldots, m_l) \tag{21}$$

whenever $j \geqslant l$ *and* $k_1 \geqslant m_1, \ldots, k_l \geqslant m_l$.

Proof. Note first that for a finite set M, $x_i \in \mathbb{N}_0$ for $i \in M$, $n \in \mathbb{N}$ and $k_1,\dots,k_n \in \mathbb{N}_0$ the inequality

$$\sum_{\substack{i_1,\dots,i_n \in M \\ \text{all distinct}}} (x_{i_1})_{k_1} \cdots (x_{i_n})_{k_n} \leqslant (s)_k \tag{22}$$

holds, where $s := \sum_{i \in M} x_i$ and $k := k_1 + \cdots + k_n$.

Fix $j \geqslant l$ and $k_1,\dots,k_j$, $m_1,\dots,m_l$ such that $k_1 \geqslant m_1,\dots,k_l \geqslant m_l$. Define $k := k_1 + \cdots + k_j$ and $m := m_1 + \cdots + m_l$. Note that $j \leqslant k$ and $l \leqslant m$. Now

$$\begin{aligned}
&\sum_{\substack{i_1,\dots,i_j=1 \\ \text{all distinct}}}^{N} (v_{i_1})_{k_1} \cdots (v_{i_j})_{k_j} \\
&= \sum_{\substack{i_1,\dots,i_l=1 \\ \text{all distinct}}}^{N} (v_{i_1})_{k_1} \cdots (v_{i_l})_{k_l} \sum_{\substack{i_{l+1},\dots,i_j=1 \\ i_{l+1},\dots,i_j \notin \{i_1,\dots,i_l\} \\ \text{all distinct}}}^{N} (v_{i_{l+1}})_{k_{l+1}} \cdots (v_{i_j})_{k_j} \\
&\overset{(22)}{\leqslant} \sum_{\substack{i_1,\dots,i_l=1 \\ \text{all distinct}}}^{N} (v_{i_1})_{k_1} \cdots (v_{i_l})_{k_l} (v_1 + \cdots + v_N - v_{i_1} - \cdots - v_{i_l})_{k_{l+1}+\cdots+k_j} \\
&= \sum_{\substack{i_1,\dots,i_l=1 \\ \text{all distinct}}}^{N} (v_{i_1})_{m_1}(v_{i_1} - m_1)_{k_1-m_1} \cdots (v_{i_l})_{m_l}(v_{i_l} - m_l)_{k_l-m_l} \\
&\quad \times (N - v_{i_1} - \cdots - v_{i_l})_{k_{l+1}+\cdots+k_j} \\
&\leqslant (N-m)_{k-m} \sum_{\substack{i_1,\dots,i_l=1 \\ \text{all distinct}}}^{N} (v_{i_1})_{m_1} \cdots (v_{i_l})_{m_l},
\end{aligned}$$

as

$$(v_{i_1} - m_1)_{k_1-m_1} \cdots (v_{i_l} - m_l)_{k_l-m_l}(N - v_{i_1} - \cdots - v_{i_l})_{k_{l+1}+\cdots+k_j} \leqslant (N-m)_{k-m}$$

whenever $v_{i_1} \geqslant m_1,\dots,v_{i_l} \geqslant m_l$, which follows from the general formula $(x_1)_{k_1} \cdots (x_r)_{k_r} \leqslant (x_1 + \cdots + x_r)_{k_1+\cdots+k_r}$, $r \in \mathbb{N}$, $x_1,\dots,x_r,k_1,\dots,k_r \in \mathbb{N}_0$. Together with the equation $(N)_m(N-m)_{k-m} = (N)_k$ this leads to

$$\frac{1}{(N)_k} \sum_{\substack{i_1,\dots,i_j=1 \\ \text{all distinct}}}^{N} (v_{i_1})_{k_1} \cdots (v_{i_j})_{k_j} \leqslant \frac{1}{(N)_m} \sum_{\substack{i_1,\dots,i_l=1 \\ \text{all distinct}}}^{N} (v_{i_1})_{m_1} \cdots (v_{i_l})_{m_l}. \tag{23}$$

Now take expectations and use the exchangeability of the offsping variables $v_1,\dots,v_N$ to verify that $\Phi_j(k_1,\dots,k_j) \leqslant \Phi_l(m_1,\dots,m_l)$. □

Remark. Note that Lemma 3.1 with $l = 1$ and $m_1 = 2$ leads to

$$\limsup_{N\to\infty} \frac{\Phi_j^{(N)}(k_1,\dots,k_j)}{c_N} \leqslant 1$$

whenever at least one of the $k_1,\dots,k_j$ is greater than or equal to 2.

For $j,k_1,\dots,k_j\in\mathbb{N}$ define

$$\phi_j(k_1,\dots,k_j):=\lim_{N\to\infty}\frac{\Phi_j^{(N)}(k_1,\dots,k_j)}{c_N}=\lim_{N\to\infty}\frac{E((v_1)_{k_1}\cdots(v_j)_{k_j})}{N^{k_1+\cdots+k_j-j}c_N}, \tag{24}$$

whenever the limit exists. The proof of the following lemma is given in Möhle and Sagitov (1999a).

Lemma 3.2. *Fix $j\in\mathbb{N}$ and $k_1,\dots,k_j\geqslant 2$. The limit* (24) *exists if and only if the limit* (11) *exists and in this case these two limits are equal.*

Remark. In particular, the lemma shows that $\psi_1(3)=0$ if and only if $\phi_1(3)=0$, i.e. that the conditions (5) and (6) are equivalent. Note that the lemma is only valid if all the $k_1,\dots,k_j$ are greater or equal than 2. In general the equation $\phi_j(k_1,\dots,k_j)=\psi_j(k_1,\dots,k_j)$ is not satisfied if some of the $k_1,\dots,k_j$ are equal to 1.

Theorem 3.3. *The process* (3) *converges in $D_E([0,\infty))$ to the Wright–Fisher diffusion with generator* (2) *if and only if the condition* (5) (*or equivalently* (6)) *is satisfied.*

Proof. Assume first that the condition (5) or equivalently (6) is satisfied, i.e. $\phi_1(3)=0$. A similar proof as presented for Lemma 5.5 in Möhle and Sagitov (1999b) shows that this condition implies that $\lim_{N\to\infty}c_N=0$ and also that $\phi_2(2,2):=\lim_{N\to\infty}\Phi_2(2,2)/c_N=0$. From Lemma 3.1 it follows that $\phi_j(k_1,\dots,k_j)=0$ for all $k_1,\dots,k_j\geqslant 2$ except for the case when $j=1$ and $k_1=2$ where we have $\phi_1(2)=1$. Thus by Lemma 3.2 the same holds for the limits $\psi_j(k_1,\dots,k_j)$ whenever $k_1,\dots,k_j\geqslant 2$. The recursion (9) yields now that all the $\psi_j(k_1,\dots,k_j)$ $(k_1,\dots,k_j\in\mathbb{N})$ are equal to zero except $\psi_1(2)=1$ and $\psi_2(1,1)=-1$. Now conclude from (15) that $a_k(x)=0$ for all $k\geqslant 3$. Thus the generator (18) has the form (2). Now apply Theorem 2.3(b).

In order to prove the converse of the theorem note that for $i\in E_N=\{0,1/N,\dots,1\}$ and with the notation $x:=i/N$ it is already known from the second remark after Lemma 2.1 that $E(C_i/N-x)=0$, $E((C_i/N-x)^2)=x(1-x)c_N$ and $E((C_i/N-x)^3)=x(1-x)(1-2x)\Psi_1(3)$. Choosing $f(x):=x^3$ it follows that

$$\begin{aligned}T_Nf(x)-f(x)&=\sum_{k=1}^{3}\frac{f^{(k)}(x)}{k!}E\left(\left(\frac{C_i}{N}-x\right)^k\right)\\&=3x^2(1-x)c_N+x(1-x)(1-2x)\Psi_1(3)\end{aligned}$$

and $Af(x)=\frac{1}{2}x(1-x)f''(x)=3x^2(1-x)$ leads to

$$\frac{T_Nf(x)-f(x)}{c_N}-Af(x)=x(1-x)(1-2x)\frac{\Psi_1(3)}{c_N}.$$

By assumption the process $(\mathscr{X}_{[t/c_N]}/N)_{t\geqslant 0}$ converges in $D_E([0,\infty))$ to the Wright–Fisher diffusion with generator (2), i.e. (19) is satisfied (in particular for the polynomial

$f(x)=x^3$ considered here). Thus we have

$$0=\lim_{N\to\infty}\sup_{x\in E_N}\left|x(1-x)(1-2x)\frac{\Psi_1(3)}{c_N}\right|$$

and hence $\psi_1(3)=\lim_{N\to\infty}\Psi_1(3)/c_N=0$, i.e. the condition (5) is satisfied. □

4. Looking backward in time and the concept of duality

The results in this section are based on coalescent theory going back to Kingman (1982a, b, c) and further developed by many authors (see Donnelly and Tavaré, 1995; Hudson, 1991; Li and Fu, 1999; Möhle, 2000a for some reviews). The coalescent theory has been extended to coalescents with multiple collisions (Pitman, 1999; Sagitov, 1999) and recently to coalescents with simultaneous multiple mergers of ancestral lines (Möhle and Sagitov, 1999a, b; Schweinsberg, 2000a, b). Furthermore the concept of duality (Liggett, 1985) is used to couple the forward process $\mathscr{X}$ with the ancestral process $\mathscr{D}:=(\mathscr{D}_r)_{r\in\mathbb{N}_0}$ arising when considering the population backward in time. By definition $\mathscr{D}_r$ counts the number of ancestors r generations backward in time. Recall (Möhle, 1999) that the process $\mathscr{X}$ is dual in the sense of Liggett (1985) to the process $\mathscr{D}$ with respect to the function H_N given by

$$H_N(i,k):=\frac{\binom{i}{k}}{\binom{N}{k}}=\prod_{j=0}^{k-1}\frac{i-j}{N-j}=\prod_{j=0}^{N-i-1}\frac{N-k-j}{N-j}, \tag{25}$$

i.e.

$$E(H_N(\mathscr{X}_r,k)\,|\,\mathscr{X}_0=i)=E(H_N(i,\mathscr{D}_r)\,|\,\mathscr{D}_0=k), \tag{26}$$

where $r\in\mathbb{N}_0$ and $i,k\in\{0,\ldots,N\}$. Assume now that the conditions of Theorem 2.3 are satisfied. We consider the case $c:=\lim_{N\to\infty}c_N=0$. The case $c>0$ is studied in a similar way. From the theory recently developed for the coalescent with simultaneous and multiple mergers (Möhle and Sagitov, 1999a) it follows that the time-scaled ancestral process $(\mathscr{D}_{[t/c_N]})_{t\geqslant 0}$ converges in the Skorohod sense to a death process $D:=(D_t)_{t\geqslant 0}$ as N tends to infinity. Based on the transition rates known for the coalescent with simultaneous and multiple mergers it is straightforward to verify that the rates of the process D are given by

$$\lim_{h\searrow 0}\frac{P(D_{t+h}=j\,|\,D_t=k)}{h}=\frac{k!}{j!}\sum_{\substack{k_1,\ldots,k_j\in\mathbb{N}\\ k_1+\cdots+k_j=k}}\frac{\phi_j(k_1,\ldots,k_j)}{k_1!\cdots k_j!}, \tag{27}$$

$j,k\in\mathbb{N}$ with $j<k$. Note (see for example Möhle, 2000b) that the Kingman coalescent appears in the limit, i.e. that D is a pure death process with rates $d_k=k(k-1)/2$ if and only if the condition (5) (or equivalently (6)) is satisfied, i.e. if and only if X is the Wright–Fisher diffusion with generator (2).

Choosing $r:=[t/c_N]$ with $t\geqslant 0$ and assuming that i/N converges to some $x\in E$ as N tends to infinity we conclude from (26) that the limit process X and the limit death process D satisfy the duality relation

$$E(X_t^k\,|\,X_0=x)=E(x^{D_t}\,|\,D_0=k). \tag{28}$$

Shortly written this means $E^x(X_t^k) = E^k(x^{D_t})$ for all $t \geqslant 0$, $x \in E$ and $k \in \mathbb{N}_0$. The same formula (28) is valid for the case $c > 0$ replacing $t \geqslant 0$ by $r \in \mathbb{N}_0$. Thus X is dual to D with respect to the measurable bounded function H on $E \times \mathbb{N}_0$ given by $H(x,k) := x^k$. In particular $E^x(X_t) = x$ and $\mathrm{Var}^x(X_t) = (1-\mathrm{e}^{-t})x(1-x)$ as $P(D_t = 1 \mid D_0 = 2) = 1 - \mathrm{e}^{-t}$.

As an application of the duality relation (28) it is shown in the following lemma how the joint moments of X are related to the distribution of the death process D. For convenience only the case $c = 0$ (when the limit processes are time-continuous) is considered. Similar results in which joint moments are expressed in terms of the dual process have been obtained for Fleming–Viot superprocesses, an infinite-dimensional analogue of the Wright–Fisher diffusion (Ethier and Krone, 1995, Lemma 2.1).

Lemma 4.1. *For $x \in E$, $m \in \mathbb{N}$, $t_1, \ldots, t_m \in \mathbb{R}$ with $0 \leqslant t_1 < \cdots < t_m$ and $k_1, \ldots, k_m \in \mathbb{N}_0$ we have*

$$E^x(X_{t_1}^{k_1} \cdots X_{t_m}^{k_m}) = \sum_{l_m=0}^{k_m + l_{m+1}} \sum_{l_{m-1}=0}^{k_{m-1}+l_m} \cdots \sum_{l_1=0}^{k_1+l_2} x^{l_1} \prod_{j=1}^{m} P(D_{t_j} = l_j \mid D_{t_{j-1}} = k_j + l_{j+1}),$$

where $t_0 := 0$ and $l_{m+1} := 0$ and $E^x := E(. \mid X_0 = x)$ denotes the expectation given the process X starts in $X_0 = x$.

Proof. Induction on m. For $m = 1$ the formula is equivalent to

$$E^x(X_{t_1}^{k_1}) = \sum_{l_1=0}^{k_1} x^{l_1} P(D_{t_1} = l_1 \mid D_0 = k_1),$$

which follows directly from the duality relation (28). To verify the step from m to $m+1$ note that

$$\begin{aligned} E^x(X_{t_1}^{k_1} \cdots X_{t_{m+1}}^{k_{m+1}}) &= \int_{E^{m+1}} x_1^{k_1} \cdots x_{m+1}^{k_{m+1}} P^x(X_{t_1} \in \mathrm{d}x_1, \ldots, X_{t_{m+1}} \in \mathrm{d}x_{m+1}) \\ &= \int_{E^m} x_1^{k_1} \cdots x_m^{k_m} P^x(X_{t_1} \in \mathrm{d}x_1, \ldots, X_{t_m} \in \mathrm{d}x_m) \\ &\quad \times \int_E x_{m+1}^{k_{m+1}} P^{x_m}(X_{t_{m+1}-t_m} \in \mathrm{d}x_{m+1}). \end{aligned}$$

Using

$$\begin{aligned} \int_{E^m} x_{m+1}^{k_{m+1}} P^{x_m}(X_{t_{m+1}-t_m} \in \mathrm{d}x_{m+1}) &= E^{x_m}(X_{t_{m+1}-t_m}^{k_{m+1}}) \stackrel{(28)}{=} E^{k_{m+1}}(x_m^{D_{t_{m+1}-t_m}}) \\ &= \sum_{l_{m+1}=0}^{k_{m+1}} x_m^{l_{m+1}} P(D_{t_{m+1}-t_m} = l_{m+1} \mid D_0 = k_{m+1}) \\ &= \sum_{l_{m+1}=0}^{k_{m+1}} x_m^{l_{m+1}} P(D_{t_{m+1}} = l_{m+1} \mid D_{t_m} = k_{m+1}) \end{aligned}$$

we conclude that

$$
\begin{aligned}
&E^x(X_{t_1}^{k_1}\cdots X_{t_{m+1}}^{k_{m+1}})\\
&\quad=\sum_{l_{m+1}=0}^{k_{m+1}} P(D_{t_{m+1}}=l_{m+1}\,|\,D_{t_m}=k_{m+1})\\
&\qquad\times\int_{E^m} x_1^{k_1}\cdots x_{m-1}^{k_{m-1}}x_m^{k_m+l_{m+1}}\,P^x(X_{t_1}\in \mathrm{d}x_1,\ldots,X_{t_m}\in \mathrm{d}x_m)\\
&\quad=\sum_{l_{m+1}=0}^{k_{m+1}} P(D_{t_{m+1}}=l_{m+1}\,|\,D_{t_m}=k_{m+1})\,E^x(X_{t_1}^{k_1}\cdots X_{t_{m-1}}^{k_{m-1}}X_{t_m}^{k_m+l_{m+1}})
\end{aligned}
$$

and by induction the lemma is established. □

Remark. The previous lemma ensures that the finite-dimensional distributions of the death process D determine those of the forward process X and vice versa. The following commutative diagram summarizes the results. The arrows are explained below.

	forward in time		backward in time
		(i)	
finite population size	$\boxed{\text{descendant process } \mathcal{X}}$	$\longleftrightarrow$	$\boxed{\text{ancestral process } \mathcal{D}}$
	(ii) $\downarrow$		$\downarrow$ (iii)
infinite population size	$\boxed{\text{limit process } X}$	$\longleftrightarrow$	$\boxed{\text{death process } D}$
		(iv)	

(i) duality of $\mathscr{X}$ and $\mathscr{D}$ with respect to the function $H_N(i,k) := \binom{i}{k} / \binom{N}{k}$,
(ii) weak convergence in $D_{[0,1]}([0,\infty))$,
(iii) weak convergence in $D_{\mathbb{N}_0}([0,\infty))$,
(iv) duality of X and D with respect to the function $H(x,k) := x^k$.

5. Examples

Models such that condition (5) is satisfied, i.e. such that the limit process is the Wright–Fisher diffusion are well known. For example for the classical Wright–Fisher model, where it is assumed that the family size vector $(v_1,\dots,v_N)$ has a symmetrical multinomial distribution, it follows that $\Phi_j(k_1,\dots,k_j)=(N)_jN^{-k}$. Thus $\Phi_1(3)/c_N=1/N$ and hence $\psi_1(3)=\phi_1(3)=0$, i.e. the condition (5) is satisfied. Even simpler is the situation for the Moran model where it is assumed that one (randomly chosen) individual has exactly two offspring, another (randomly chosen) individual has no offspring and all the other $N-2$ individuals have exactly one offspring. In this case the coalescence probability is given by $c_N=1/\binom{N}{2}=2/(N(N-1))$. From $v_1\leqslant 2$ almost surely it follows immediately that $\Phi_1(3)=0$. In particular (5) is satisfied and Theorem 3.3 is applicable.

We now focus on models where the assumptions of Theorem 2.3 are satisfied but not the condition (5), i.e. where the limit process X exists but is not equal to the Wright–Fisher diffusion.

Assume that with some probability $q\in(0,1)$ each individual has exactly one offspring and with probability $p:=1-q$ each individual has either N offspring (with probability $1/N$) or no offspring (with probability $1-1/N$). The coalescence probability for this model is p. As $v_1+\cdots+v_i$ takes the values 0, i, N with probabilities $p(1-i/N)$, q and pi/N, respectively, we conclude that

$$a_k(x)=(-x)^k(1-x)+(1-x)^kx=\int_E(y-x)^k\,v(x,\mathrm{d}y)$$

with $v(x,\Gamma):=(1-x)\varepsilon_0(\Gamma)+x\varepsilon_1(\Gamma)$, where ε_y denotes the point measure in $y\in E$. The generator (18) is therefore given by

$$\begin{aligned}Af(x)&=\sum_{k=1}^{\infty}\frac{f^{(k)}(x)}{k!}a_k(x)\\&=(1-x)\sum_{k=1}^{\infty}\frac{f^{(k)}(x)}{k!}(0-x)^k+x\sum_{k=1}^{\infty}\frac{f^{(k)}(x)}{k!}(1-x)^k\\&=(1-x)(f(0)-f(x))+x(f(1)-f(x))\\&=\int_E(f(y)-f(x))\,v(x,\mathrm{d}y),\end{aligned}$$

which follows also from $T_Nf(x)=f(0)p(1-x)+f(x)q+f(1)px$, i.e.

$$\frac{T_Nf(x)-f(x)}{p}=(1-x)(f(0)-f(x))+x(f(1)-f(x)).$$

Note that if p does not depend on N we are in the situation (a) of Theorem 2.3 with $\mu(x,.)=p(1-x)\varepsilon_0+q\varepsilon_x+px\varepsilon_1$. If $p=p_N$ depends on N such that p_N converges to zero as N tends to infinity then we are in the situation (b) of Theorem 2.3.

Finally a model similar to the example in Möhle and Sagitov (1999a) is discussed. Fix a constant $l\in\mathbb{N}$ and assume for convenience that N is a multiple of l. Consider such an exchangeable population model where exactly l individuals have N/l offspring while all the other $N-l$ family sizes are zero. In this case $(v_1+\cdots+v_i)/N$ takes the value j/l with the hypergeometric probability $\binom{l}{j}(i)_j(N-i)_{l-j}/(N)_l$.

Choosing $i=[Nx]$ with $x\in E$ this probability converges to the Binomial probability $B(l,x)(j):=\binom{l}{j}x^j(1-x)^{l-j}$ as N tends to infinity, i.e. the measure $\mu_N(x,.)$ defined in the second remark after Lemma 2.1 converges weakly to the measure $\mu(x,.)$ given by $\mu(x,j/l):=B(l,x)(j)$, $j\in\{0,\dots,l\}$. From $E((v_1)_2)=(N/l)_2\, l/N=(N-l)/l$ it follows that the coalescence probability is given by $c_N=(N-l)/(l(N-1))$ which converges to $c:=l^{-1}>0$ as N tends to infinity. The corresponding polynomials $a_k(x)$ have the form

$$a_k(x)=\frac{1}{c}\int_E (y-x)^k\mu(x,\mathrm{d}y)=\ l\sum_{j=0}^{l}\left(\frac{j}{l}-x\right)^k\binom{l}{j}x^j(1-x)^{l-j}$$

and the generator (20) is given by

$$Af(x)=\frac{1}{c}\int_E (f(y)-f(x))\mu(x,\mathrm{d}y)=\ l\sum_{j=0}^{l}f\left(\frac{j}{l}\right)\binom{l}{j}x^j(1-x)^{l-j}-l f(x).$$

Acknowledgements

The author would like to thank Frank Baldus for helpful discussions concerning the generator (18) of the limit process X.

References

Cannings, C., 1974. The latent roots of certain Markov chains arising in genetics: a new approach, I. Haploid models. Adv. Appl. Prob. 6, 260–290.

Cannings, C., 1975. The latent roots of certain Markov chains arising in genetics: a new approach, II. Further haploid models. Adv. Appl. Prob. 7, 264–282.

Crow, J.F., Kimura, M., 1970. An Introduction to Population Genetics Theory. Harper & Row, New York.

Donnelly, P., Tavaré, S., 1995. Coalescents and genealogical structure under neutrality. Annu. Rev. Genet. 29, 401–421.

Ethier, S.N., Krone, S.M., 1995. Comparing Fleming–Viot and Dawson–Watanabe processes. Stoch. Proc. Appl. 60, 171–190.

Ethier, S.N., Kurtz, T.G., 1986. Markov Processes. Characterization and Convergence. Wiley, New York.

Ethier, S.N., Nagylaki, T., 1980. Diffusion approximations of Markov chains with two time scales and applications to population genetics. Adv. Appl. Prob. 12, 14–49.

Feller, W., 1971. An Introduction to Probability Theory and Its Applications, Vol. I, 2nd Edition. Wiley, New York.

Fisher, R.A., 1922. On the dominance ratio. Proc. Roy. Soc. Edin. 42, 321–431.

Hudson, R.R., 1991. Gene genealogies and the coalescent process. Oxf. Surv. Evol. Biol. 7, 1–44.

Kingman, J.F.C., 1982a. On the genealogy of large populations. J. Appl. Prob. 19 A, 27–43.

Kingman, J.F.C., 1982b. Exchangeability and the evolution of large populations. In: Koch, G., Spizzichino, F. (Eds.), Exchangeability in Probability and Statistics. North–Holland Publishing Company, Amsterdam, pp. 97–112.

Kingman, J.F.C., 1982c. The coalescent. Stoch. Process. Appl. 13, 235–248.

Li, W.H., Fu, Y.X., 1999. Coalescent theory and its application in population genetics. in: Halloran, M.E., Geisser, S. (Eds.), Statistics in Genetics, IMA Volumes in Mathematics and Its Applications, Vol. 112. Springer, New York.

Liggett, T.M., 1985. Interacting Particle Systems. Springer, Berlin.

Möhle, M., 1998. Robustness results for the coalescent. J. Appl. Prob. 35, 438–447.

Möhle, M., 1999. The concept of duality and applications to Markov processes arising in neutral population genetics models. Bernoulli 5, 761–777.

Möhle, M., 2000a. Ancestral processes in population genetics—the coalescent. J. Theor. Biol. 204, 629–638.
Möhle, M., 2000b. Total variation distances and rates of convergence for ancestral coalescent processes in exchangeable population models. Adv. Appl. Prob. 32, 983–993.
Möhle, M., Sagitov, S., 1999a. A Classification of Coalescent Processes for Haploid Exchangeable Population Models. Chalmers University of Technology, Department of Mathematics, Preprint No. 10. Ann. of Probab., accepted for publication.
Möhle, M., Sagitov, S., 1999b. Coalescent Patterns in Diploid Exchangeable Population Models. Berichte zur Stochastik und verwandten Gebieten, Johannes Gutenberg—Universität Mainz, July 1999, ISSN 0177-0098, J. Math. Biol., submitted for publication.
Norman, M.F., 1972. Markov Processes and Learning Models. Academic, New York.
Pitman, J., 1999. Coalescents with multiple collisions. Ann. Probab. 27, 1870–1902.
Sagitov, S., 1999. The general coalescent with asynchronous mergers of ancestral lines. J. Appl. Prob. 36, 1116–1125.
Sato, K., 1976. Diffusion processes and a class of Markov chains related to population genetics. Osaka J. Math. 13, 631–659.
Schweinsberg, S., 2000a. A necessary and sufficient condition for the Λ-coalescent to come down from infinity. Electron. Comm. Probab. 5, 1–11.
Schweinsberg, S., 2000b. Coalescents with simultaneous multiple collisions. Electron. J. Probab. 5, 1–50.
Trotter, H.F., 1958. Approximations of semi-groups of operators. Pacific J. Math. 8, 887–919.
Wright, S., 1931. Evolution in Mendelian populations. Genetics 16, 97–159.